Bayesian Data Analytics for Reliability Modeling Improvement

Mingyang Li
Department of Industrial and Management Systems Engineering
University of South Florida
Jan 26th, 2018
Research Expertise and Highlights:

- Multi-level Data Integration and Analytics for Mission Critical System Reliability Assessment, Testing, Diagnosis, Prognosis and Real-Time Health Management.

- Multi-fidelity Data Integration and Analytics for Crowd Surveillance Improvement.

- Various Data Science and System Informatics Methods Development and Diverse Applications.

Some of Current and Former Team Members in DSSI Lab at Dept. of IMSE:

[membership images]

Quality & Reliability  Healthcare  HVAC  Wind energy  Water
Outline

• Background

• Part I - Multi-level Data Fusion

• Part II - Heterogeneous Data Quantification

• Summary
• Focus:
  Bayesian Data Analytics for Reliability Modeling Improvement
Key Word: Bayesian

- Parameter Learning

Classic Statistics

Data $\rightarrow$ Parameters

Limited Data or No Data

Bayesian Statistics

Data $\rightarrow$ Posterior

Parameter Learning

Flexible & Coherent

Methodology I: Multi-level Data Fusion

- External data sources
- Domain knowledge
- Non-informative prior

Prior
Key Word: Bayesian (Cont’d)

• Model Learning

Classic Statistics

- Model 1
  Data → Parameters

- Model m
  Data → Parameters

• Underfitting/Overfitting
• Inefficient

Bayesian Statistics

- Data → Posterior

- Parameters & Model

- Para. Prior

- Model Prior

Efficient & Effective

Methodology II: Heterogeneity Quantification
Key Word: Reliability Modeling

- Reliability: product quality over time\cite{1}
  \[ \Pr(T > t) \]
  Time-to-failure

- Reliability modeling

- Data Feature
  - Non-negative and asymmetric
  - Covariates
  - Censoring
  - Others: availability, heterogeneity, etc.
Lifecycle View of Reliability Modeling

Marketing

Requirements

Maintenance

Design and Development

Reliability Modeling

- Warranty data
- Warranty policy
- Product Req.
- Specifications
- Functional Relationship
- Evaluation, allocation, etc.
- Test data
- Test plans, etc.
- Assessment
- Production information
- Production, etc.
- Burn-in, etc.
- Repair Logs
- Maintenance policy

Production

Testing
Part I - Multi-level Data Fusion:

Bayesian Multi-level Information Aggregation for Hierarchical Systems Reliability Modeling Improvement
Vision

Heating Ventilating & Air-Conditioning (HVAC) System\(^2\)

Crowd Surveillance System\(^4\)

Data-rich Environment:
Data Fusion

EEG/MEG (high-temporal-resolution)\(^3\)
fMRI (high-spatial-resolution)\(^3\)
Focus: System Reliability

- **Performance index**: system reliability

- **Modeling Challenges**:
  - Expensive system-level tests
  - Scarce/absent engineering knowledge
  - Complex failure relationship
  - High requirement on reliability assessment

- **Research Goal**:
  Improve system-level reliability modeling by utilizing all reliability information throughout the system in a **systematic and coherent** manner.
Opportunity I: Hierarchical System Structure

Electro-Mechanical-Actuator (EMA) System

Power Supply (PS)  Actuator Servo Drive (ASD)  DC Motor

Elements in System Hierarchy

- EMA System
  - PS Sub-system
    - Motor PS
  - Logic PS
  - Controller
  - ASD Sub-system
    - Bridge
  - DC Motor

Divide & Conquer
Opportunity II: Multi-source Multi-level Data

- **Multi-source reliability information**: prior knowledge (e.g., domain knowledge, historical studies, etc.) + ongoing reliability test data.

<table>
<thead>
<tr>
<th>Elements</th>
<th>Prior knowledge</th>
<th>Reliability Test Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower-level</td>
<td>Familiar</td>
<td>(1) Abundant (2) Limited but easy to collect</td>
</tr>
<tr>
<td>Upper-level</td>
<td>Unfamiliar or unknown</td>
<td>(1) Absent (2) Limited and/or expensive/hard to collect</td>
</tr>
</tbody>
</table>

- **Multi-level information imbalance**

  ![Diagram showing multi-level information imbalance]

  **Aggregation**

  **Reliability Information:**
  - Prior knowledge
  - Reliability test data
  - Absent information
State of the Art

<table>
<thead>
<tr>
<th>Methodology Summary</th>
<th>System Reliability Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Parametric methods</strong></td>
</tr>
<tr>
<td>Multi-level information aggregation</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

- **Features of the proposed model:**
  - Failure-time data with covariates and censoring
  - Semi-parametric modeling
  - Information aggregation from lower levels
Overview of the Proposed Work

Upward Recursively

1. Data of $X_{(l-1,1)}$
2. Aggregated posterior of $X_{(l-1,1)}$
3. Combined Prior of $X_{(l-1,1)}$
4. Posterior of $X_{(l-1,1)}$

Native Prior of $X_{(l-1,1)}$
Induced Prior of $X_{(l-1,1)}$
Modeling of Individual Element

Proportion Hazard Model (PHM):

\[ H_{(l,k)}(t) = H^b_{(l,k)}(t) \exp(\beta^T_{(l,k)} u_{(l,k)}) \]

Baseline cumulative hazard function
Cumulative hazard function
Baseline cumulative hazard increment
Covariate coefficient
Covariate

Gamma Process Prior:
\[ Z(t) \approx G(c_0 \alpha(t), c_0) \]
Confidence parameter
Mean function

Multivariate Normal Prior:
\[ \pi(\beta_{(l,k)}) = N(\mu_{(l,k)}, \Sigma_{(l,k)}) \]
Mean vector
Covariance matrix

Baseline cumulative hazard increments:
\[ \Delta H^b_{j,(l,k)} \sim Gamma(c_0(\alpha(s_j) - \alpha(s_{j-1})), c_0) \]
Carry information for aggregation

Unique Reliability:
\[ R_{(l,k)}(t) \leftrightarrow H_{(l,k)}(t) \]
Aggregation Procedure: Step 1

- **Step 1** – Compute the posterior (lower-level element):

\[ p(\beta_{(l,k)}, \Delta H^b_{j,(l,k)} \mid \Omega_{(l,k)}) \propto L(\beta_{(l,k)}, \Delta H^b_{j(l,k)} \mid \Omega_{(l,k)}) \pi(\beta_{(l,k)}) \pi(\Delta H^b_{j(l,k)}) \]

Joint Posterior \( \propto \) Likelihoods \( \times \) Joint Priors:

Integration of data and prior

Failure-time data with covariates and censoring

Example:

\[ 0 \rightarrow t_5 \rightarrow t_1 \rightarrow t_3 \rightarrow t_4 \rightarrow t_2 \rightarrow t_3 \]

\( t_n \): the actual failure time stamp of test unit \( n \)

Bayesian PHM integrates the reliability prior knowledge and failure data
Aggregation Procedure: Failure Relationship

- Failure relationship between two levels:

  General Relationships

  Reliability functions:
  \[ R_{(l-1,k)}(t) = f\left(R_{(l,k)}(t)\right), \ \kappa \in Q_{(l-1,k)} \]

  Baseline cumulative hazard increments:
  \[ A\Delta H^b_{j,(l-1,k)} = g\left(\Delta H^b_{j,(l,k)}\right), \ \kappa \in Q_{(l-1,k)} \]

  Example: Series configuration

  Aggregated
  \[ A R_{(l-1,1)}(t) = R_{(l,1)}(t) R_{(l,2)}(t) \]

  \[ A\Delta H^b_{j,(l-1,1)}(t) = \Delta H^b_{j,(l,1)}(t) + \Delta H^b_{j,(l,2)}(t) \]

  Information is aggregated through \( \Delta H^b_{j} \) based on failure relationships
**Aggregation Procedure: Steps 2-3**

- **Step 2** - Aggregate the posterior:

  Aggregated posterior: \( A \Delta H_{j,(l-1,1)}^b = g(\Delta H_{j,(l,\kappa)}^b), \kappa = 1, 2 \)

- **Step 3** - Approximate the induced prior:

  Induced prior: \( I \Delta H_{j,(l-1,1)}^b \leftarrow A \Delta H_{j,(l-1,1)}^b \)

  \( I \Delta H_{j,(l-1,1)}^b \sim \text{Gamma}(I \eta_{j,(l-1,1)}, I \lambda_{j,(l-1,1)}) \)

  (Validate by K-S goodness fitness test)
Aggregation Procedure: Step 4

• **Step 4** – Combine the native prior and the induced prior:

  Combined prior: \( ^C \Delta H_{j,(l-1,1)}^b \sim \text{Gamma}(^C \eta_{j,(l-1,1)}, ^C \lambda_{j,(l-1,1)}) \)

  \[
  ^C \eta_{j,(l-1,1)} = w^I \eta_{j,(l-1,1)} + (1-w)^N \eta_{j,(l-1,1)}
  \]

  \[
  ^C \lambda_{j,(l-1,1)} = w^I \lambda_{j,(l-1,1)} + (1-w)^N \lambda_{j,(l-1,1)}
  \]

  \( w \): balance native prior and induced prior

  Weighting factor: \( 0 \leq w \leq 1 \)

• **Step 1** – Compute the posterior (higher-level element):

Similar Bayesian inference

\( ^C \Delta H_{j,(l-1,1)}^b \sim \text{Gamma}(^C \eta_{j,(l-1,1)}, ^C \lambda_{j,(l-1,1)}) \)
Information Aggregation: Procedure Review

- Recursive
- Flexible
- Generic

Aggregate Upward Recursively

Step 1

Posterior: \( \Delta H_{j,(l-1,1)}^b \)

Combined prior: \( C \Delta H_{j,(l-1,1)}^b \)

Step 2

Aggregated posterior: \( A \Delta H_{j,(l-1,1)}^b \)

Level \( l - 1 \)

Level \( l \)

Failure data: \( \Omega_{(l-1,1)} \)

Induced prior: \( I \Delta H_{j,(l-1,1)}^b \)

Native prior: \( N \Delta H_{j,(l-1,1)}^b \)

Step 3

Induced prior: \( I \Delta H_{j,(l-1,1)}^b \)

Native prior: \( N \Delta H_{j,(l-1,1)}^b \)

Combined prior: \( C \Delta H_{j,(l-1,1)}^b \)

Failure data: \( \Omega_{(l-2,1)} \)

Failure data: \( \Omega_{(l,1)} \)

Failure data: \( \Omega_{(l,2)} \)

Posterior: \( A \Delta H_{j,(l,1)}^b \)

Posterior: \( A \Delta H_{j,(l,2)}^b \)

Posterior: \( A \Delta H_{j,(l,2)}^b \)

Failure data: \( \Omega_{(l,2)} \)

Failure data: \( \Omega_{(l,2)} \)

prior: \( C \Delta H_{j,(l,1)}^b \)

prior: \( C \Delta H_{j,(l,1)}^b \)

prior: \( C \Delta H_{j,(l,2)}^b \)
Numerical Case Study

- A two-level hierarchical system with 3 elements
- One covariate $u$ is considered with binary values: 0/1
- Test data are simulated with 30 intervals
Information Aggregation

- Steps 1-2: Compute and aggregate the posteriors of components:
  \[ A \Delta H_{j,(l-1,k)}^b = g(\Delta H_{j,(l,\kappa)}^b), \kappa \in Q_{(l-1,k)} \]

- Step 3: Approximate the aggregated posteriors into the induced priors:

<table>
<thead>
<tr>
<th>Variable ((i,1))</th>
<th>(\Delta H_{j,(1,1)}^{b})</th>
<th>(\Delta H_{j,(1,2)}^{b})</th>
<th>(\Delta H_{j,(2,1)}^{b})</th>
<th>(\Delta H_{j,(2,2)}^{b})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\bar{\kappa} )</td>
<td>27.4611</td>
<td>26.7808</td>
<td>29.0416</td>
<td>2.8181</td>
</tr>
<tr>
<td>(\alpha )</td>
<td>107.8274</td>
<td>93.0126</td>
<td>69.9600</td>
<td>10.1614</td>
</tr>
<tr>
<td>(\beta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\gamma )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\delta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\varepsilon )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\zeta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\eta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\theta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\iota )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\kappa )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\xi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\omicron )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\pi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\rho )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\tau )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\upsilon )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\phi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\chi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\psi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\omega )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\varepsilon )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\theta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\iota )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\kappa )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\xi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\omicron )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\pi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\rho )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\tau )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\upsilon )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\phi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\chi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\psi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\omega )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\varepsilon )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\theta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\iota )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\kappa )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\xi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\omicron )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\pi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\rho )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\tau )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\upsilon )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\phi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\chi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\psi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\omega )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\varepsilon )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\theta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\iota )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\kappa )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\xi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\omicron )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\pi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\rho )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\tau )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\upsilon )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\phi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\chi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\psi )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\omega )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\varepsilon )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\theta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\iota )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\kappa )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• Step 4: Combine the induced priors and the native priors

\[ w=0, 0.2, 0.8 \] Different effects of information aggregation

• Step 1: Compute posteriors for the system

Posterior comparison of hazard increment at the 5th interval for \( X_{(1,1)} \)

**Series:**

- \( w=0 \), posterior mean 0.668
- \( w=0.2 \), posterior mean 0.47
- \( w=0.8 \), posterior mean 0.251

**Parallel:**

- \( w=0 \), posterior mean 0.211
- \( w=0.2 \), posterior mean 0.123
- \( w=0.8 \), posterior mean 0.091
Series System Reliability Curve Comparisons

Reliability Curve Comparison for the Series System

- **TRUE**
- **w=0**
- **w=0.2**
- **w=0.8**
- **Frequentist**
Parallel System Reliability Curve Comparisons

Reliability Curve Comparison for the Parallel System

- **TRUE**
- **w=0**
- **w=0.2**
- **w=0.8**
- **Frequentist**

$R(t)$ vs. Time Stamp $t$
Part II - Heterogeneous Data Quantification:

Bayesian Modeling and Learning of Heterogeneous Time-to-Event Data with an Unknown Number of Sub-populations
**Health Care Utilization Data**[^19]

- **Frequency (%)**
  - Bar chart showing frequency distribution of visits to hospital.
  - Excessive zeros at zero visits.

**Nanocrystals Growth Data**[^21]

- Picture of nanocrystals with size indication.

**Alloy Fatigue Crack Size Data**[^20]

- Line graph showing crack size (inches) vs. millions of cycles.
  - Failure threshold indicated.

**Heterogeneous Populations:**

- Heterogeneous Data Quantification
  - How many?
  - How to model?
Focus: Time-to-Event Data

- Time-to-event (TTE) data is important

TTE: Time to occurrence of an event of interest

- Machine breakdown
- Contract renew
- Surgery completion
- Product recall
- Occurrence of a disease

Event

GENERAL & CRITICAL
TTE Heterogeneity

- TTE: assembly time

Intelligent Robotic Assembly System[22]

- Homogenous assumption [Crossed out]

- Reason: heterogeneous products quality, etc.
TTE Heterogeneity (Cont’d)

- Reliability examples
- **Semiconductor industry**\(^{[23]}\): infant mortality failures
  *Reason*: manufacturing defects, assembly errors, etc.
- **Automobile industry**\(^{[24]}\): early failures
  *Reason*: material quality, unverified design changes, etc.
- **Industry with evolving technology**\(^{[25]}\): heterogeneity especially critical
  *Reason*: immature technology

Q: How to model TTE heterogeneity?
Heterogeneity Modeling of TTE

- **Change point model**\(^{[26-28]}\):
  - \( h(t) \)
  - Data
    - Model 1
    - Model 2
  - Limitation: different domains

- **Frailty model**\(^{[29,30]}\):
  - Sub-populations labels 1, 2, 3, ...
  - Limitation: known membership

- **Mixture model**\(^{[31,32]}\):
  - \( h(t) \)
  - Data
    - Model 1
  - Scope
    - (1) One model under entire domain
    - (2) Unknown membership
    - (3) Meaningful interpretation;
    - (4) Feedback information.
## Mixture Model: Gaps and Solutions

<table>
<thead>
<tr>
<th>Existing Method</th>
<th>Limitation</th>
<th>Advantage</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Known</strong> the number of sub-populations $(m)$[^33,^34]</td>
<td>subjective</td>
<td><strong>Unknown</strong> $m$, learned from data</td>
<td>objective</td>
</tr>
<tr>
<td>Model estimation + model selection (e.g., LRT, AIC)[^35,^36]</td>
<td>Two-step</td>
<td>Bayesian formulation</td>
<td>Joint model estimation and selection</td>
</tr>
<tr>
<td>Mixtures of distributions[^32,^37,^38,^41]</td>
<td>w/o covariates</td>
<td>Mixtures of regressions</td>
<td>w/ covariates</td>
</tr>
<tr>
<td><strong>Conjugate</strong> prior[^39,^40]</td>
<td>Restrictive</td>
<td><strong>Non-conjugate</strong> prior</td>
<td>Generic</td>
</tr>
</tbody>
</table>
• Assuming an **unknown** # of sub-populations

• Considering influence of possible **covariates**

• Achieving **joint** model estimation and model selection

• Comprehensive treatment of **non-conjugate** priors
Mixture Model: Known $m$

- $j^{th}$ homogenous sub-population:
  \[ h_j(t \mid x) = h_j^b(t) \exp(\beta_j^T x) \]
  - Hazard function
  - Baseline Hazard function
  - Covariates
  - Covariate coefficients
  - Unique

**Benefits:** (1) Covariates; (2) Flexible; (3) $h_j(t) \leftrightarrow f_j(t)$

- The overall heterogeneous population:
  \[ g(t \mid \Theta^m, x) = \sum_{j=1}^m w_j f_j(t \mid \theta_j, x) \]
  - Population pdf
  - Sub-population pdf
  - All unknowns
  - Sub-population proportion
  - Sub-population unknowns
  - What if unknown
  - $m$ known

M. Li
Mixture Model: Unknown $m$

- Finite mixture model:
  \[ g(t \mid \Theta^m, x) = \sum_{j=1}^{m} w_j f_j(t \mid \theta_j, x) \]

$m$ choices

\[ f_1 \quad f_2 \quad \ldots \ldots \quad f_m \]

\[ t_i \]

Infinite choices

\[ f_1 \quad f_2 \quad f_3 \quad \ldots \ldots \]

\[ t_i \]

Solution: Dirichlet Process
Mixture Model: Unknown $m$ (Cont’d)

- Bayesian hierarchical formulation:

\[
\begin{align*}
  t \mid x, \theta &\sim f(\cdot \mid x, \theta), \\
  \theta \mid P &\sim P \\
  P \mid \alpha, P_0 &\sim \text{DP}(\alpha P_0(\cdot))
\end{align*}
\]

\[
g(t \mid \Theta, x) = \sum_{j=1}^{\infty} w_j f_j (t \mid \theta_j, x)
\]

A random distribution \hspace{1cm} Dirichlet process \hspace{1cm} Base distribution

Positive scalar

finite mixture: \hspace{1cm} New formulation:

\[
\sum_{j=1}^{m} w_j f_j (t \mid \theta_j, x)
\]

(1) no restriction on $m$

(2) $m$ learned objectively

infinite mixture: \hspace{1cm} (3) Joint model estimation and model selection

\[
\sum_{j=1}^{\infty} w_j f_j (t \mid \theta_j, x)
\]
Estimation Challenges

Data: \( \mathbf{D} = \{t_i, \Delta_i, \mathbf{x}_i\}_{i=1}^n \)

Right-censored indicator

Unknowns: \( \Theta = \{w_j, \beta_j, k_j, \eta_j\}_{j=1}^\infty \)

Weibull baseline, shape, scale

Joint posterior:
\[
\pi(\Theta | \mathbf{D}) \propto \prod_{i=1}^n \left( \sum_{j=1}^\infty w_j f_j \left(t_i \mid \beta_j, k_j, \eta_j, \mathbf{x}_i \right) \right)^{\Delta_i} \cdot \left( \sum_{j=1}^\infty w_j R_j \left(t_i \mid \beta_j, k_j, \eta_j, \mathbf{x}_i \right) \right)^{1-\Delta_i} \cdot \pi(\Theta)
\]

Challenges:

1. High dependency
2. Non-conjugate prior
3. Infinite # of unknowns

- Slow/failed convergence
- Sampling difficulty
- Computationally formidable
1. High dependency: \( Z = \{ z_i \}_{i=1}^n \)
   labels for \( t_i \)’s
   \[
   \pi(\eta_j | \cdot) \quad \text{M-H} \quad \text{Reparameterization: } \lambda_j = \eta_j^{-k_j} \]
   \[
   \pi(k_j | \cdot) \quad \text{M-H} \quad \text{ARS, condition provided} \]
   \[
   \pi(\beta_j | \cdot) \quad \text{M-H} \quad \text{ARS, condition provided} \]

2. Non-conjugate prior: \( \pi(\beta_j | \cdot) \) \( \pi(\eta_j | \cdot) \)

**Metropolis-Hasting (M-H)**\(^{[41]}\):
- Pros: General purpose
- Cons: Tuning problem, samples auto-correlated

**Adaptive Rejection Sampling (ARS)**\(^{[42]}\):
- Condition-based
- No turning, samples independent

3. Infinite # of unknowns: slice-sampling techniques\(^{[40]}\): \( j=1,2,...,J^* \), where \( J^* \) is finite
Realized Features of the Proposed Work

- **Unknown** # of sub-populations: Dirichlet process
- **Covariates**: hazard regression
- **Joint** model estimation & selection: Bayesian model
- **Non-conjugate** priors: a series of sampling techniques
Numerical Case Study: Effectiveness

• **Simulation setup**
  - 2-mixture of Weibull regression
  - Single covariate \( X \sim \text{Unif}(0,5) \)
  - Right-censored time \( 1.0 \times 10^5 \)

## Table 1. Model estimation results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sub-population 1</th>
<th>Sub-population 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p_1 )</td>
<td>( k_1 )</td>
</tr>
<tr>
<td>True value</td>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Estimate</td>
<td>0.28</td>
<td>0.66</td>
</tr>
</tbody>
</table>
Efficiency

- **M–H only**: Convergence failed
  - $m^* = 2$

- **M–H with reparameterization**: Converged
  - $m^* = 2$

- **ARS with reparameterization**: Fast convergence
  - $m^* = 2$
Real data analysis

- Assembly time data

(a) Estimated densities comparison
(b) UTP curves comparison

**Figure 4.** Comparisons of models w/ and w/o considering heterogeneity
System Informatics & Data Analytics

Diverse Applications

Data View
- Volume
- Variety
- Velocity
- Veracity

System View
- Planning
- Prognosis
- Diagnosis
- Monitoring
- Prediction
- Testing
- Modeling

Practice
- Control
- Method
- Quality & Reliability

HVAC
Combustion
Wind energy
Crowd Surveillance
Water
Healthcare
Solar
Nanotechnology
Thanks 😊


